

**UNITED STATES PATENT APPLICATION**  
**FOR**  
**SYSTEMS AND METHODS FOR BROWSING MULTIMEDIA CONTENT**  
**ON SMALL MOBILE DEVICES**

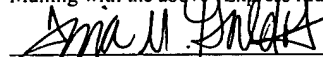
Inventors:

Patrick Chiu  
Maryam Kamvar  
Tohru Fuse  
Surapong Lertisithichai  
Sandeep Casi  
Lynn D. Wilcox

**CERTIFICATE OF MAILING BY "EXPRESS MAIL"**  
**UNDER 37 C.F.R. §1.10**

"Express Mail" mailing label number: EV 375096306 US  
Date of Mailing: March 31, 2004

I hereby certify that this correspondence is being deposited with the United States Postal Service, utilizing the "Express Mail Post Office to Addressee" service addressed to **Mail Stop PATENT APPLICATION, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450**, and mailed on the above Date of Mailing with the above "Express Mail" mailing label number.

  
\_\_\_\_\_  
Tina Galdos  
Signature Date: March 31, 2004

# **SYSTEMS AND METHODS FOR BROWSING MULTIMEDIA CONTENT ON SMALL MOBILE DEVICES**

## **Inventors**

Patrick Chiu  
Maryam Kamvar  
Tohru Fuse  
Surapong Lertisithichai  
Sandeep Casi  
Lynn D Wilcox

## **COPYRIGHT NOTICE**

[0001] A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

## **FIELD OF THE DISCLOSURE**

[0002] The present disclosure relates to the storing, processing, and browsing of multimedia data.

## **BACKGROUND**

[0003] Current advances in mobile and wireless technology are making it easier to access multimedia contents anywhere and anytime. A multimedia content can include, but is not limited to, a video, a video segment, a keyframe, an image, a graph, a figure, a drawing, a picture, a text, a keyword, and other suitable contents. The cutting edge technology provides the possibility to watch multimedia contents on a small mobile device, which can be, but is not limited to, a PDA, a cell phone, a Tablet PC, a Pocket PC, and other suitable electronic devices. The small mobile device can utilize an associated input device such as a pen or a stylus to interact with a user. However, it is challenging to browse

multimedia content on the small mobile device for a number of reasons. First, the small screen area of such device restricts the amount of multimedia content that can be displayed; Second, user interaction tends to be more tedious on the small mobile device, and the limited responsiveness of the current generation of such devices is another source of aggravation; Third, due to bandwidth and performance issues, it is necessary to carefully select the portions of the multimedia content to transmit over a network. Furthermore, despite the high portability and flexibility of the small mobile devices serving as mobile multimedia terminals, how they handle and process multimedia contents huge in term of number of bytes generally is a big challenge, because the resources of these small mobile devices are potentially limited.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0004] **Figure 1** is an illustration of an exemplary multimedia content browsing system in accordance with one embodiment of the invention.

[0005] **Figure 2** is a flow chart illustrating an exemplary multimedia content browsing process in accordance with one embodiment of the invention.

[0006] **Figure 3 (a)-(c)** are illustrations of a multimedia content composed from other multimedia contents in accordance with one embodiment of the invention.

[0007] **Figure 4 (a)-(c)** are illustrations of exemplary content layers in accordance with one embodiment of the invention.

[0008] **Figure 5 (a)-(c)** are illustrations of keywords associated with a multimedia content in accordance with one embodiment of the invention.

[0009] **Figure 6 (a)-(c)** are illustrations of an exemplary widget layer in accordance with one embodiment of the invention.

[0010] **Figure 7** is an illustration of the composition of keyframes from two multimedia contents in accordance with one embodiment of the invention.

[0011] **Figure 8 (a)-(c)** are illustrations of exemplary configurations of scalable architectures in accordance with one embodiment of the invention.

[0012] **Figure 9** is an illustration of an exemplary multimedia content browsing system in accordance with one embodiment of the invention.

## DETAILED DESCRIPTION

[0013] The invention is illustrated by way of example and not by way of limitation in the figures of the accompanying drawings in which like references indicate similar elements. It should be noted that references to “an” or “one” or “some” embodiment(s) in this disclosure are not necessarily to the same embodiment, and such references mean at least one.

[0014] Systems and methods in accordance with the present invention enable the browsing of multimedia contents on small mobile devices. They smoothly blend three key tasks of multimedia content browsing: querying the multimedia contents by keywords, exploring the search results by viewing keyframes of the multimedia contents, and playing a stream of the multimedia contents, e.g., videos or video segments. During each task, only the necessary portions (e.g., titles, keyframes, video segments) of the multimedia contents are retrieved and rendered, thereby putting less demand on a communication network, which can be, but is not limited to, Internet, an intranet, a local area network, a wireless network, a Bluetooth network, and other suitable concepts. Videos can be stored in a segment-based multimedia content database, which is designed to support the browsing, retrieval, storage and reuse of multimedia contents, such as videos. A layered imaging model is introduced in order to browse the multimedia contents effectively on the small screen area, and as a way to transition between tasks. Each layer may have its own transparent value set individually, continuously, and interactively, and the layers can overlap on top of each other when rendered on the screen.

[0015] Since a small mobile device alone may not have enough resources to handle the entire task of multimedia content browsing, a scalable architecture can be adopted to break up the task using the small mobile device as a browsing component, a Hard Disk Drive (HDD) hosting a multimedia content database, and a resource-rich computing device as a processing component. Here, the resource-rich computing device can include, but is not limited to, a desktop PC, a laptop PC, a workstation, a server and a mainframe computer; the HDD can be, but is not

limited to one of: an external HDD, a portable HDD, a wireless HDD, a Bluetooth HDD, and an internal HDD on a resource-rich computing device.

[0016] The application software used by the multimedia content browsing system can be implemented in Java, wherein Java2D is used to support the rendering of the layers, and QuickTime for Java is used to play the stream of contents.

[0017] **Figure 1** is an illustration of an exemplary system in an embodiment of the present invention. Although this diagram depicts objects/processes as logically separate, such depiction is merely for illustrative purposes. It will be apparent to those skilled in the art that the objects/processes portrayed in this figure can be arbitrarily combined or divided into separate software, firmware and/or hardware components. Furthermore, it will also be apparent to those skilled in the art that such objects/processes, regardless of how they are combined or divided, can execute on the same computing device or can be distributed among different computing devices connected by one or more networks or other suitable communication means.

[0018] Within the exemplary multimedia content browsing system **100** in **Figure 1**, a browsing component **101** is capable of rendering one or more layers **102** of multimedia contents, such as videos or video segments, on a screen **104**. The transparency values of each of the one or more layers can be set interactively using a widget layer **103**, which is operable via one or more input devices **105**. The browsing component communicates with a processing component **107** via a communication network **106**. During the query task, a search engine component **109** in the processing component retrieves multimedia contents from a multimedia content database **110**, such contents may contains keyframes and metadata of the text and keywords associated with the contents. During the exploration and/or content playing task, a content composition component **108** transmits the multimedia contents back to the browsing component for rendering, after necessary composition, animation, and storage of the contents.

[0019] **Figure 2** is a flow chart illustrating an exemplary multimedia content browsing process in accordance with one embodiment of the invention. Although this figure depicts functional steps in a particular order for purposes of

illustration, the process is not limited to any particular order or arrangement of steps. One skilled in the art will appreciate that the various steps portrayed in this figure could be omitted, rearranged, combined and/or adapted in various ways.

[0020] Referring to **Figure 2**, the query from a user requesting multimedia contents is processed at step **201**. The multimedia contents are then retrieved from the multimedia content database and explored at step **202**. At step **203**, the transparency values are set for multiple layers of contents, e.g. title list, and keyframes. These contents can then be composed and animated at step **204**, if necessary, using the set transparency values at step **203**. These composed contents can be stored in the multimedia content database before being transmitted and rendered for display at step **205**.

[0021] In some embodiments, the multimedia content database is designed to support the retrieval of a video by keyword query. In a typical database, keywords are associated with the video as a whole. In some embodiments, for example, if keywords are obtained from a time-aligned translation, keywords may be associated with particular timestamps, which are actually part of the metadata associated with the video (each multimedia content in the multimedia content database has at least one timestamp, allowing multimedia contents such as images or texts to serve as indexes or links into other multimedia contents, e.g., videos). Keyword-based video retrieval from this type of database results in a list of relevant videos, with optional marks showing where keywords occur in each of the videos.

[0022] In some embodiments, the multimedia content database is segment-based to support browsing, retrieval, and reuse of a video by segments. The video is first segmented, either manually or using any standard automatic video segmentation algorithm. Keywords are associated with each of these video segments, either by manually annotating the segments or by associating time-stamped keywords with the corresponding segments. A keyword query results in a list of relevant segments. Relevance is determined based on the number of occurrences of the keyword in the segment, where the number is possibly weighted. A relevance score for an entire video is computed as the sum of the relevance scores of all of

its component segments. Thus a keyword query can result in an ordered list of all relevant videos, with information on the relevance of the segments in each of the videos.

[0023] In some embodiments, the segment-based multimedia content database is also designed to support reuse of video segments. Using an editing application, users can create a video by concatenating segments from multiple source videos retrieved from the database. The database keeps track of the source of a video where segments from multiple videos in the database are re-used to create the video. The graphical representation of the presence of a source video associated with a video segment 301 is indicated with a black, downward-pointing arrow, as shown in **Figure 3 (a)**. Gesturing down (**Figure 3 (b)**) on the video segment expands the view to show the source video of that segment, shown as 302 in **Figure 3 (c)**, and gesturing up collapses the view. The composed video is stored back into the multimedia content database, and its segments contain links to the source video that the segment was taken from.

[0024] In some embodiments, the browsing component provides effective features to browse multiple segments of videos on the screen of a small mobile device. These features are based on a layered image model, comprising one or more content layers and a widget layer on the browsing component to set the gradient transparency values of the layers. Two or more content layers can be overlapped on top of each other, and the tasks of query, exploration, and playing of segmented multimedia contents can be switched smoothly among each other. By considering the whole process as an integration of several tasks rather than a bunch of isolated tasks, the browsing component makes the interaction with users less haphazard and more fluid.

[0025] In some embodiments, the browsing component supports the query of, e.g., segmented videos, by keywords. A simple text box, in a style used by a search engine such as Google™, accepts keywords from the input devices and searches the segmented videos in the segment-based multimedia content database. The search results are displayed as a list of video titles on a first content layer, which can be visible in any opaque color, as shown in **Figure 4 (a)**. On the left

margin of each title is a bar whose height indicates the relevance score for that video.

[0026] In some embodiments, the browsing component supports the task of exploring the search results generated by the keyword query. Such exploration involves checking the promising query results in the list and looking at their keyframes by selecting videos in the list one at a time. The selected video is highlighted in red, and a second content layer, which is transparent, appears showing a keyframe from the selected video, as shown in **Figure 4 (b)**. This keyframe can be one of the first frame of the video, the last frame of the video, and the most characteristic frame of the video that is extracted by a standard video segmentation algorithm. Notice that the first content layer now becomes transparent and on top of the second content layer, and both are visible.

[0027] In some embodiments, the transparency values of the first content layer showing the query results and the second content layer showing the keyframe are automatically changed to make it possible to see both layers when they are overlaid on top of each other during the transitioning from query to exploration. The transparency value of the first layer drops from 1.0 to alpha 1, and the transparency value of the second layer rises from 0.0 to alpha 2, where the effect of  $\alpha 1 = \alpha 2 = 0.8$  is shown in **Figure 4 (b)**.

[0028] In some embodiments, the transparency values of the two layers can be adjusted manually and in continuous gradient values. Sometimes it is desirable to adjust the transparency values to see better either the query result on the first content layer or the keyframe on the second content layer. For a small mobile device used under different lighting conditions, having greater visual separation between the layers may be more helpful than just uniformly changing the display's brightness or contrast.

[0029] In some embodiments, the segments of the video are visualized by a graphical representation at the bottom part of the screen, and each segment has a bar whose height indicates its relevance. The keywords associated with the selected video **501** are shown at the bottom of the screen above the segment bars in **Figure 5 (a)**. The user can scroll the set of keywords by gesturing left or right



on top of the keywords, as shown in **Figure 5 (b)**, and the result **502** after scrolling is shown in **Figure 5 (c)**.

[0030] In some embodiments, the browsing component supports the playing of a stream of contents, when some interesting video segments are found. There are several ways to play the video:

- Tapping on a selected video plays that video from the beginning;
- Tapping on a segment of a video plays the video from the first frame of that segment;
- Tapping on a segment in the expanded segment view plays from that source segment's first frame.

[0031] The transition from exploration to content playing is made smooth by playing the video "in-place" of the keyframe, and hiding the query results. The visual effect is that the second content layer switches from a transparent keyframe layer to become an opaque video layer, and the first content layer is faded out. As the result, only the second content layer is visible, as shown in **Figure 4 (c)**.

[0032] In some embodiments, a small video controller is activated on the top edge of the video layer, and users can stop, pause, or jump to another video segment on the time slider. Users can also tap on the segments on the bottom of the screen to jump to another part of the video.

[0033] In some embodiments, the browsing component accepts a gesture made via an input device such as a stylus anywhere over a content layer to adjust its transparency value continuously. If the stylus is held down before gesturing, the widget layer showing transparent gradient appears, as shown in **Figure 6 (a)**, where the dot **601** shows the current transparency values ( $x$  = first content layer,  $y$  = second content layer). Gesturing to the right (**Figure 6 (b)**) decreases the transparency of the first content layer and gesturing to the left increases it. Similarly, gesturing up decreases the transparency of the second content layer and gesturing down increases it. The result after the gesturing is shown in **Figure 6 (c)**, where the dot **602** shows the current transparency values of the two content layers.

[0034] In some embodiments, the processing component is capable of carrying out the heavy duty tasks of image processing, such as the searching and retrieving of multimedia contents from the multimedia content database, the composition and/or animation of multiple layers of contents, e.g. keyframes, using the transparency values of the layers set by the user. **Figure 7** illustrates the composition of keyframes in accordance with some embodiments of the present invention. Here keyframes are extracted from video segments A and B, respectively, and the transparent ratios between the transparency values of content layers showing keyframes from A and B are set as 25%, 50%, and 75%, respectively. The content composition component generates the composed and/or animated images for each of these three keyframes at each of the three transparent ratios, using any standard image composition and/or animation algorithm.

[0035] In some embodiments, the composed contents of the three keyframes of a video or a video segment are also stored in the multimedia content database, in addition to the actual video or video segment. If the number of videos or video segments under exploration is  $N$ , for example, then at least  $3*N(N-1)/2$  of the composed contents should be generated. The composed contents can also be transmitted over the communication network and rendered on a content layer on the browsing component.

[0036] In some embodiments, the task of browsing multimedia contents using the multimedia content database, the browsing component, and the processing component has to be distributed among multiple computing devices using a scalable architecture. A small mobile device, which is often used as the browsing component, usually does not have enough storage to store multimedia contents such as videos on its own body. A large amount of Compact Flash memory or a Micro Drive would increase the storage area of the small mobile device, but it's still not enough to handle a large volume of multimedia contents. In addition, the small mobile device also lacks processing power to compose keyframes of contents and generating animations. Besides, the multimedia content database usually runs on an HDD in a high performance server placed in a data center. Therefore, if a small mobile device needs to browse a video that is stored in the

multimedia content database, or it needs to view composed images of keyframes generated dynamically, it needs to access one or more computing devices remotely through a communication network. On the other hand, recent high performance laptop PC has enough power to compose images and generate animation in real time and is often used as the processing component, which traditionally has to be a desktop PC, a workstation, a server, or a mainframe computer. Therefore, the scalable architecture can be constructed with, for example, the combination of a wireless HDD, a desktop PC, a portable laptop PC, a communication network, and a small mobile device. Three types of configurations of the scalable architecture to handle the browsing of multimedia contents on a small device are described as Pocket, Portable, and Network.

[0037] In some embodiments, the Pocket configuration is adopted, which comprises a wireless (Bluetooth) HDD 802 to host the multimedia content database, a desktop PC 803 as the processing component, and a small mobile device like a PDA 801 as the browsing component, as shown in **Figure 8 (a)**. Such configuration enables making the small mobile device to watch multimedia contents in handheld style. The desktop PC composes keyframe images for the video, and then generates animation that consists of the composed images. The wireless HDD stores the actual multimedia contents, and also the composed contents that the desktop PC previously generated. A small mobile device is capable of playing a stream of video content through a wireless network, such as Bluetooth.

[0038] In some embodiments, the Portable configuration is adopted, which comprises a high performance laptop PC 804 as the processing component, the HDD of the high performance laptop PC to host the multimedia content database, and a small mobile device like a PDA as the browsing component, as shown in **Figure 8 (b)**. The laptop PC generates the composed and animated contents in real time on each request from PDA dynamically, as shown in **Figure 9**.

[0039] In some embodiments, the Network configuration is adopted, which in addition to the Portable configuration, further comprises a server 805 that is placed at a data center to host the multimedia content database, as shown in

**Figure 8 (c).** The small mobile device can communicate with the laptop PC through a wireless network like Bluetooth, and the laptop PC can communicate with the server in the data center through Internet.

[0040] One embodiment may be implemented using a conventional general purpose or a specialized digital computer or microprocessor(s) programmed according to the teachings of the present disclosure, as will be apparent to those skilled in the computer art. Appropriate software coding can readily be prepared by skilled programmers based on the teachings of the present disclosure, as will be apparent to those skilled in the software art. The invention may also be implemented by the preparation of integrated circuits or by interconnecting an appropriate network of conventional component circuits, as will be readily apparent to those skilled in the art.

[0041] One embodiment includes a computer program product which is a machine readable medium (media) having instructions stored thereon/in which can be used to program one or more computing devices to perform any of the features presented herein. The machine readable medium can include, but is not limited to, one or more types of disks including floppy disks, optical discs, DVD, CD-ROMs, micro drive, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, DRAMs, VRAMs, flash memory devices, magnetic or optical cards, nanosystems (including molecular memory ICs), or any type of media or device suitable for storing instructions and/or data.

[0042] Stored on any one of the computer readable medium (media), the present invention includes software for controlling both the hardware of the general purpose/specialized computer or microprocessor, and for enabling the computer or microprocessor to interact with a human user or other mechanism utilizing the results of the present invention. Such software may include, but is not limited to, device drivers, operating systems, execution environments/containers, and applications.

[0043] The foregoing description of the preferred embodiments of the present invention has been provided for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise forms

disclosed. Many modifications and variations will be apparent to the practitioner skilled in the art. Particularly, while the concept “keyframe” is used in the embodiments of the systems and methods described above, it will be evident that such concept can be interchangeably used with equivalent concepts such as image, and other suitable concepts. Embodiments were chosen and described in order to best describe the principles of the invention and its practical application, thereby enabling others skilled in the art to understand the invention, the various embodiments and with various modifications that are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the following claims and their equivalents.